**Rate Limits for OpenAI**

Rate limits are measured in five ways: RPM (requests per minute), RPD (requests per day), TPM (tokens per minute), TPD (tokens per day), and IPM (images per minute). Rate limits can be hit across any of the options depending on what occurs first. For example, you might send 20 requests with only 100 tokens to the ChatCompletions endpoint and that would fill your limit (if your RPM was 20), even if you did not send 150k tokens (if your TPM limit was 150k) within those 20 requests.

A screenshot of a website

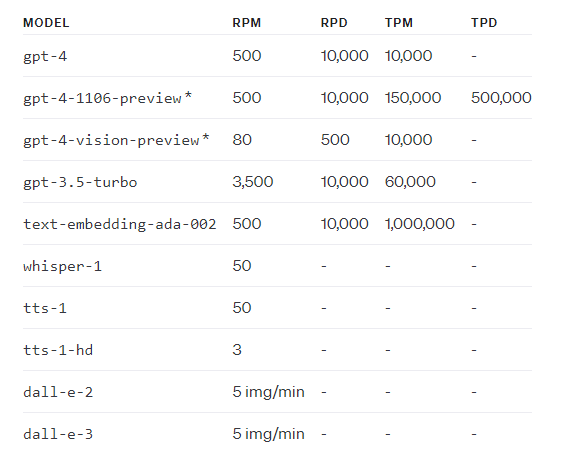
Description automatically generated

**Free :**

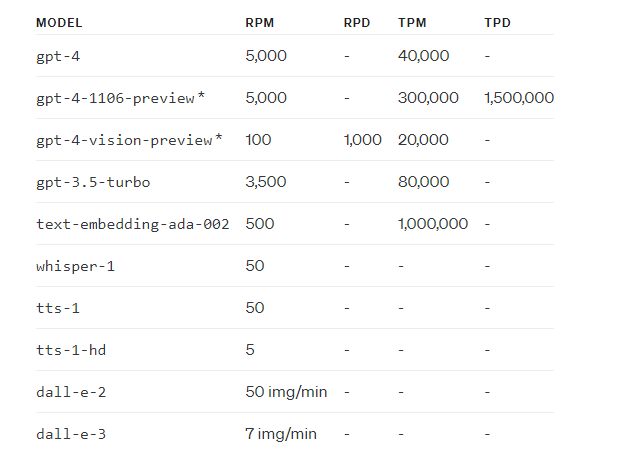
A screenshot of a computer

Description automatically generated

Tier 1



Tier 2



Tier 3

A screenshot of a computer

Description automatically generated

Tier 4

A screenshot of a computer

Description automatically generated

Tier 5

A screenshot of a computer

Description automatically generated